



The number of distinct symbols in sections of rectangular arrays

Sherman K. Stein^a, Sándor Szabó^b

^a*Department of Mathematics, University of California at Davis, 1 Shields Avenue, Davis, CA 95616-8633, USA*

^b*Institute of Mathematics and Informatics, University of Pécs, Ifjúság u. 6, 7624 Pécs, Hungary*

Received 27 October 2003; received in revised form 11 October 2005; accepted 31 October 2005

Available online 6 January 2006

Abstract

We investigate transversals of rectangular arrays. For positive integers m and n , where $2 \leq m \leq n$ an m by n array consists of mn cells arranged in m rows and n columns. Each cell contains one symbol. When $m = n$ we speak of an array of order n . A *section* in the array consists of m cells, one from each row and no two from the same column. A *transversal* is a section whose m symbols are distinct. A *partial transversal* is a subset of a transversal. We investigate the existence in an array of a section with many different symbols, in particular the existence of a transversal.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Transversal; Latin square; Array

1. Introduction

For over two centuries, going back to Euler in 1782, the study of transversals was restricted to a special type of array, a Latin square. In an n by n Latin square the symbols $1, 2, \dots, n$ appear in each row and in each column. Euler observed that if a Latin square does not have a transversal, then it has no orthogonal mate. He also showed that the group table of a cyclic group of even order has no transversal [6, pp. 302–303].

Ryser [14] conjectured that every Latin square of odd order has a transversal, and, more generally, that the number of transversals of a Latin square has the same parity as the order of the square. However, Parker [13] pointed out that many Latin squares of order 7 have an even number of transversals, for instance, (6) and many other cases in [11]. Confirming half of Ryser's conjecture, Balasubramanian [2] proved that a Latin square of even order has an even number of transversals.

It is trivial that a Latin square of order n has a section with at least $n/2$ distinct symbols. (Pick any cell in the first row, then a cell in the second row with a different symbol, and so on. This procedure can continue for at least $n/2$ rows. Then adjoin cells to form a section.) Various authors have obtained much stronger results: Koksma [10], Drake [4], Woolbright [19], Shor [15], Fu et al. [7] showed that the number of distinct symbols in some transversal is at least

$$n - (1/3)n, \quad n - (1/4)n, \quad n - \sqrt{n}, \quad n - (5.53)(\log n)^2, \quad n - (5.518)(\log n)^2,$$

respectively.

E-mail addresses: stein@math.ucdavis.edu (S.K. Stein), sszabo7@hotmail.com (S. Szabó).

In [17] Stein generalized the problem from Latin squares to any square array. For instance, he showed that in an array of order n where each element appears exactly n times there is a section with at least approximately $(0.63)n$ distinct elements.

Erdős and Spencer [5] showed that an array of order n in which each symbol appears at most $(n-1)/16$ times has a transversal.

It is convenient to introduce $L(m, n)$, the largest integer such that if each symbol in an m by n array appears at most $L(m, n)$ times, then the array must have a transversal. For instance, the preceding result shows that $L(n, n) \geq (n-1)/16$.

The algorithm in which you try to construct a transversal by working down row by row yields a different lower bound. If each symbol appears at most k times in an m by n array, the algorithm is certainly successful if $(m-1)k < n-1$. This implies $L(m, n) \geq (n-1)/(m-1)$.

A result of Hall [8] lends some support for our conjecture that $L(n-1, n) = n-1$. Consider an abelian group $A = \{a_1, a_2, \dots, a_n\}$ of order n and b_1, b_2, \dots, b_{n-1} , a sequence of $n-1$ elements of A , not necessarily distinct. Construct an $n-1$ by n array by placing $b_i a_j$ in the cell where row i meets column j . Hall proved that such an array has a transversal.

Snevily [16] offered a conjecture closely related to Hall's theorem: any square submatrix of the group table of an abelian group of odd order has a transversal.

2. Some values of $L(m, n)$

Note that $L(m+1, n) \leq L(m, n)$ and that $L(m, n) \leq L(m, n+1)$. The function L satisfies two more inequalities, stated in Theorems 2.1 and 2.2.

Theorem 2.1. *If $n \leq 2m-2$, then $L(m, n) \leq n-1$.*

Proof. The proof rests on a construction due to Parker [13], illustrated for the cases $(m, n) = (4, 4)$, $(4, 5)$, and $(4, 6)$:

1	1	4	4
2	2	1	1
3	3	2	2
4	4	3	3

1	1	1	4	4
2	2	2	1	1
3	3	3	2	2
4	4	4	3	3

1	1	1	4	4	4
2	2	2	1	1	1
3	3	3	2	2	2
4	4	4	3	3	3

In each case an attempt to construct a transversal might as well begin with a 1 in the top row. The 2 must then be selected from the 2s in the second row, and the 3 from the 3s in the third row. Such choices do not extend to a transversal. \square

Theorem 2.2. $L(m, n) < mn/(m-1)$.

Proof. This follows from the fact that if only $m-1$ distinct symbols appear in an m by n array, the array cannot have a transversal. In detail, if each of $m-1$ symbols appears at most k times and $(m-1)k \geq mn$, the symbols may fill all the cells. \square

Though Theorem 2.2 is valid for all n , in view of Theorem 2.1, it is of interest only for $n \geq 2m-1$. In Akbari et al. [1] prove that for $m \geq 2$ and $n \geq 2m^3 - 6m^2 + 6m - 1$, $L(m, n)$ equals $\lfloor (mn-1)/(m-1) \rfloor$. On the other hand van Rees [18] has shown that for $m \geq 2$ and $n = m^2 - 3n + 3$, $L(m, n)$ is less than $\lfloor (mn-1)/(m-1) \rfloor$.

A moment's thought shows that $L(2, 2) = 1$ and that $L(2, n) = 2n-1$ for $n \geq 3$. This means that for $m=2$ the inequalities in Theorems 2.1 and 2.2 become the best possible.

The case $m=3$ is similar, for it turns out that $L(3, n)$ equals $n-1$ for $n=3, 4$ and is the largest integer less than or equal to $(3n-1)/2$ for $n \geq 5$. The following two lemmas are used in the proof of the second assertion. In each case x stands for 1 or 2.

Lemma 2.1. *Assume that in a 3 by n array, $n \geq 4$, some symbol occurs at most three times. Then, if there is no transversal some symbol occurs at least $2n-2$ times, hence at least $3n/2$ times.*

Proof. We regard two arrangements of symbols in cells as equivalent if one can be obtained from the other by a permutation of rows, a permutation of columns, and a relabeling. There are 10 inequivalent configurations of cells occupied by a symbol that occurs at most three times. We list three of these. The reader can supply the remaining seven easily.

x		

x	x	

x		
	x	

We illustrate the argument by treating the case when 1 appears twice, in one row, as in the following diagram:

1	1				...
b	b	2	b	b	...
a	a	c	a	a	...

We may assume that 2 occurs as indicated. It follows that the cells marked a are filled with 2s. This implies that all the cells marked b are also filled with 2s, and finally that the cell marked c also contains a 2. Hence the symbol 2 appears at least $2n$ times.

In the case when a symbol occurs only once, it is not hard to show that some symbol appears at least $2n - 2$ times. In most of the other cases a symbol appears almost $3n$ times. \square

Theorem 2.3. (a) $L(3, 3) = 2$ and $L(3, 4) = 3$. (b) For $n \geq 5$, $L(3, n)$ is the greatest integer less than or equal to $(3n - 1)/2$.

Proof. Exhaustive computer calculations show that

$$L(3, 3) = 2, \quad L(3, 4) = 3, \quad L(3, 5) = 7.$$

An induction shows that for even $n \geq 6$, $L(3, n) = (3n - 2)/2$ and that for odd $n \geq 5$, $L(3, n) = (3n - 1)/2$.

Assume that the induction holds for a particular even n , that is, $L(3, n) = (3n - 2)/2$. We will show that it holds for $n + 1$, which is odd, that is, $L(3, n + 1) = (3n + 2)/2$. Note that in this case we would have $L(3, n + 1) = L(3, n) + 2$.

Consider a 3 by $n + 1$ array in which each symbol occurs at most $(3n + 2)/2$ times. If each symbol occurs at most $(3n - 2)/2$ times, delete one column, obtaining a 3 by n array, which has a transversal, by the inductive assumption. Hence the original array has, also.

Now assume that there is at least one symbol occurring at least $3n/2$ times. If there are two such symbols, they occupy at least $3n$ cells. Hence some symbol appears at most three times. By Lemma 2.1 some symbol occurs at least $3(n + 1)/2$ times, which contradicts the assumption that each symbol occurs at most $(3n + 2)/2$ times.

Hence there is only one symbol that occurs at least $3n/2$ times, that is, $3n/2$ or $(3n + 2)/2$ times. There must be a column in which it appears at least twice. Deleting that column, we obtain a 3 by n array in which each symbol occurs at most $(3n - 2)/2$ times. By the inductive assumption, this array has a transversal, hence the original array does.

The argument when n is odd is similar. In this case $L(3, n + 1) = L(3, n) + 1$. This completes the proof of the theorem. \square

The next theorem gives a non-trivial lower bound on $L(m, n)$.

Theorem 2.4. $L(m, n) \geq n - m + 1$.

Proof. The proof is an induction on m .

The theorem is true for $m = 2$ or $m = 3$. Assuming it is true for $m - 1$, we will prove it for any array A with m rows. In order to simplify the diagrams and the exposition, we consider the case $m = 5$, which illustrates the argument in the general case.

Assuming that $L(4, n)$ is at least $n - 3$, we will show that $L(5, n)$ is at least $n - 4$.

Consider a 5 by n array A in which each symbol appears at most $n - 4$ times. By the induction assumption, the 4 by n array consisting of the first four rows of A has a transversal. Assuming that A does not have a transversal, we may

conclude that A contains an equivalent of the following configuration:

1					x	x	x	x	...
	2								...
		3							...
			4						...
				1	x	x	x	x	...

An x stands for 1, 2, 3, 4. There are cells marked x since we are assuming that A has no transversal.

At this point $2(n-5)+2$ cells contain x or 1. Since 1 occurs at most $n-4$ times in A , there must be a 2, 3, or 4 in some cell marked x . It is no loss of generality to take that symbol to be 2.

No matter which x is replaced by 2, there is a unique partial transversal consisting of that cell and cells marked 1, 3, and 4.

1					2	x	x	x	...
x	2					x	x	x	...
		3		x		x	x	x	...
			4						...
				1	3	x	x	x	...

Continuing the analysis, one shows that the symbols 1, 2, 3, and 4 occupy more than $4(n-4)$ cells, a contradiction. \square

3. Particular values of $L(m, n)$

In view of our experience with $m=2$ and 3, it is tempting to conjecture that the values for $L(m, n)$ suggested by Theorems 2.1 and 2.2 would be correct even for $m \geq 4$. In other words, one is tempted to conjecture that for $m \leq n \leq 2m-2$, we have $L(m, n) = n-1$ and that for $n \geq 2m-1$, we have $L(m, n)$ equal to the greatest integer less than or equal to $(mn-1)/(m-1)$. Hickerson [9] has shown that $L(4, 4) = 3$, in agreement with the first part of the conjecture. However, he also has shown by a construction and computer search that $L(4, 7)$ is 8, which is a counterexample to the second part.

The following arrays show that $L(5, 5) \leq 3$ and $L(6, 6) \leq 4$. Exhaustive computer calculations show that $L(5, 5) = 3$ and $L(6, 6) = 4$.

1	3	4	2	1
5	2	3	6	5
6	7	3	4	6
7	2	5	4	7
1	3	4	2	1

1	1	1	5	4	2
2	2	2	5	4	1
3	3	3	7	8	2
7	4	4	4	8	7
8	5	5	7	5	8
6	3	3	7	8	1

For $n \leq 7$ the following table lists the known values of $L(m, n)$. For $m=2, 3, 4$ a sudden jump occurs from $L(m, 2m-2)$ to $L(m, 2m-1)$.

	2	3	4	5	6	7
2	1	5	7	9	11	13
3		2	3	7	8	10
4			3	4	5	8
5				3	5	
6					4	

4. Other conditions

In this section we investigate two conditions that imply the existence of a section with many distinct symbols. The proof of Theorem 4.1 is a simplification due to Hickerson [9] of our original proof.

Theorem 4.1. *Consider an n by n table filled with symbols $1, 2, \dots, n/s$ such that each symbol appears exactly s times in each row and in each column, where $2 \leq s \leq n/2$. Then there is a section that contains n/s distinct symbols.*

Proof. Choose a section that contains a maximal number of distinct symbols. We may assume that this section is the main diagonal of the table since this is only a matter of rearranging rows and columns of the table. Assume that the main diagonal contains t distinct symbols. We may assume that the symbols in the first t positions in the main diagonal are distinct. In addition, we may assume that $1, 2, \dots, t$ are the symbols in the first t cells in the main diagonal since this is only a matter of exchanging the symbols. We partition the table in the following way.

A	B
C	D

Here A is a t by t table, B is a t by $n - t$ table, C is an $n - t$ by t table, D is an $n - t$ by $n - t$ table.

The main diagonal of A contains the symbols $1, 2, \dots, t$. D contains no symbol larger than t , since otherwise there is a section with more distinct symbols than the main diagonal has.

Now count the number of cells in C containing symbols greater than t . Since there are none in D, this is the same as the number in the union of C and D, that is, in the bottom $n - t$ rows. There are $n/s - t$ symbols larger than t and each occurs exactly s times in each of the $n - t$ rows. Thus, the number of cells in C containing symbols greater than t is $(n/s - t)s(n - t)$.

Next count the number of cells in the union of A and C containing symbols greater than t that is, in the left t columns. Again there are $n/s - t$ distinct symbols greater than t , and each occurs exactly s times in each of the t columns. Since C is contained in the union of A and C, we get the inequality

$$(n/s - t)s(n - t) \leq (n/s - t)st.$$

If $t < n/s$, then we can divide through by $(n/s - t)s$ to get $n - t < t$, which implies $t \geq n/2$. This contradiction shows that $t \geq n/s$. Since there are only n/s distinct symbols in the whole array, we must have $t = n/s$. \square

The proof of the next theorem illustrates the probabilistic first moment method.

Theorem 4.2. *If in an array of order n each symbol appears at most cn times, then there is a section with at least*

$$\frac{n}{c + 1 + (c - 1)/(n - 1)}$$

distinct symbols.

Proof. Consider an array of order n in which each symbol appears at most cn times. We say that the cells $[x_1, y_1]$, $[x_2, y_2]$ corresponding to the triples $[x_1, y_1, z_1]$, $[x_2, y_2, z_2]$ of the table form a special pair if

$$x_1 \neq x_2, \quad y_1 \neq y_2, \quad z_1 = z_2.$$

The cells are not in the same row and not in the same column and are filled with the same symbol. Let Γ be the set of all special pairs in the array. Let $N = \{1, 2, \dots, n\}$ and let σ be a 1–1 map from N to N . The set of cells

$$[i, \sigma(i)], \quad i \in N$$

corresponds to a section of the table.

Let Ω be the probability space whose elements are the $n!$ sections of the array. We assign the same probability to each element of Ω . Clearly $|\Omega| = n!$. For notational convenience we number the elements of Γ by $1, 2, \dots, m$. Suppose

the special pair $[x_1, y_1], [x_2, y_2]$ is numbered by i . We define a subset A_i of Ω to be all $\sigma \in \Omega$ for which

$$\sigma(x_1) = y_1, \quad \sigma(x_2) = y_2.$$

Note that

$$\Pr[A_i] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

Next we define a random variable X_i associated with A_i . Let $X_i = 1$ if $\sigma \in A_i$ and let $X_i = 0$ if $\sigma \notin A_i$. Now

$$E(X_i) = \Pr[A_i] = \frac{1}{n(n-1)}.$$

Let $X = X_1 + \cdots + X_m$. Intuitively, X is the number of special pairs that appear in a randomly chosen section

$$E(X) = \frac{m}{n(n-1)}.$$

Let k_i be the number of distinct symbols that appear exactly i times. Then

$$k_1 + 2k_2 + \cdots + cnk_{cn} = n^2.$$

The number of special pairs is at most

$$t = k_2 \binom{2}{2} + k_3 \binom{3}{2} + \cdots + k_{cn} \binom{cn}{2}.$$

So we have

$$E(X) = \frac{m}{n(n-1)} \leq \frac{t}{n(n-1)}.$$

Consider the maximum of t subject to

$$k_1 + 2k_2 + \cdots + cnk_{cn} = n^2.$$

This is an optimization problem in integer variables. However, we will switch to the corresponding problem in real variables. Since the function is linear with a simplex as a domain, its maximum occurs at a vertex. If at the vertex $(0, \dots, 0, k_i, 0, \dots, 0)$, then $ik_i = n^2$. As $i \geq 2$, it follows that

$$t = k_i \binom{i}{2} = k_i \frac{i(i-1)}{2} = n^2 \frac{i-1}{2} \leq n^2 \frac{cn-1}{2}.$$

Thus,

$$E(X) \leq n^2 \frac{cn-1}{2} \frac{1}{n(n-1)} \leq \frac{n}{2} \left[c + \frac{c-1}{n-1} \right].$$

There is a section with at least $(n/2)[c + (c-1)/(n-1)]$ special pairs.

Consider such a section. Suppose it contains r distinct symbols, say $1, 2, \dots, r$. Suppose further that symbol i appears d_i times in the section. Then

$$d_1 + d_2 + \cdots + d_r = n.$$

We want to find the minimum of r subject to

$$\binom{d_2}{2} + \cdots + \binom{d_r}{2} \leq \frac{n}{2} \left[c + \frac{c-1}{n-1} \right].$$

In short

$$d_2^2 + \cdots + d_r^2 \leq n \left[c + \frac{c-1}{n-1} + 1 \right].$$

The minimum of $d_2^2 + \cdots + d_r^2$ on a simplicial domain occurs when all d_i are equal, hence equal to $n/(r-1)$. Then we have the sum of squares equal to

$$(r-1)\left(\frac{n}{r-1}\right)^2.$$

So we wish to find r such that

$$\frac{n^2}{r-1} \leq n \left[c+1 + \frac{c-1}{n-1} \right].$$

So there are at least

$$r \geq \frac{n}{c+1 + (c-1)/(n-1)} + 1$$

distinct symbols in that section. \square

The preceding theorem is probably not the best possible. For instance, consider the case when $c=2$. Assume that each symbol in an array of order n fills the cells in two rows. Then every section has exactly $n/2$ distinct symbols. If this suggests the worst case, then we conjecture that if each symbol appears at most cn times, $c \geq 2$, then some section contains at least n/c distinct symbols.

5. Sections in multiplication tables

Paige [12] proved that the group table of an abelian group of order n has a transversal if and only if there is not exactly one element of order 2 in G . (As pointed out in [3, pp. 7–9], this theorem has a history going back to 1903.) This is the basis of the following theorem.

Theorem 5.1. *Let G be an abelian group of order n and let M be its multiplication table, an array of order n .*

- (1) *If M has a transversal, then it does not have a section with exactly $n-1$ distinct symbols.*
- (2) *If M has a section with exactly $n-1$ distinct symbols, then it does not have a transversal.*
- (3) *If n is an odd prime, then M does not have a section with exactly two distinct symbols.*

Proof. Let x_1, \dots, x_n be all the elements of G . Then M consists of the triples $[x_i, x_j, x_i x_j]$, $1 \leq i, j \leq n$.

If f is a 1–1 map from G to G , then $f(x_1), \dots, f(x_n)$ is a permutation of the elements x_1, \dots, x_n and the triples $[x_i, f(x_i), x_i f(x_i)]$, $1 \leq i \leq n$ form a transversal of M .

The elements of order two in G together with the identity element e form a subgroup H of G , which is a direct product of s cyclic groups of order two. To prove (1) assume that M has a transversal. By Paige's theorem, $s \neq 1$, that is, either $s=0$ or $s \geq 2$. Assume also that the table has a section with exactly $n-1$ distinct elements. Let the triples $[x_i, f(x_i), x_i f(x_i)]$, $1 \leq i \leq n$ form this section and its entries. The list $x_1 f(x_1), \dots, x_n f(x_n)$ contains exactly $n-1$ distinct elements of G . Thus, one element of G is missing from the list, say u , and one element of G appears twice, say v . Clearly $u \neq v$. The product $x_1 f(x_1) \cdots x_n f(x_n) u v^{-1}$ is equal to the product $x_1 \cdots x_n$. This gives $f(x_1) \cdots f(x_n) = u^{-1} v$. We know that $f(x_1), \dots, f(x_n)$ is a permutation of x_1, \dots, x_n . Every non-identity element whose order is not two can be paired with its inverse. So $f(x_1) \cdots f(x_n) = h_1 \cdots h_t$, where h_1, \dots, h_t are all the elements of H . In the case $s=0$, $H=\{e\}$ and we get the contradiction $u=v$. In the case $s \geq 2$ again $h_1 \cdots h_t = e$ and we get the contradiction $u=v$. In order to see that $h_1 \cdots h_t = e$, list 2^s sequences of 0s and 1s of length s . Each sequence forms a row of a 2^s by s table. Notice that each column contains 2^{s-1} 0s and 2^{s-1} 1s. Since the latter number is even, $h_1, \dots, h_t = e$.

The case (2) is the contrapositive of case (1).

Let us turn to the proof of (3). Suppose that n is an odd prime and M has a section with exactly two distinct elements. Suppose that the triples $[x_i, f(x_i), x_i f(x_i)]$, $1 \leq i \leq n$ form this section. There are exactly two distinct elements in the list $x_1 f(x_1), \dots, x_n f(x_n)$, say u and v . The product $x_1 f(x_1) \cdots x_n f(x_n)$ is equal to e , as is shown by pairing each element with its inverse. The product is also equal to $u^i v^{n-i}$, where $1 \leq i \leq n-1$. Thus $e = u^i v^{n-i}$. This implies $u^i = v^i$ which, since i is relatively prime to the order of G , leads to the contradiction, $u=v$. \square

References

- [1] S. Akbari, O. Etesami, H. Mahini, M. Mahmoodi, A. Sharifi, Latin transversals in long rectangular arrays, in the special issue of the *Discrete Mathematics* devoted to the 2003 IPM Workshop on Combinatorics, Linear Algebra, and Graph Coloring, to appear.
- [2] K. Balasubramanian, On transversals of latin squares, *Linear Algebra Appl.* 131 (1990) 125–129.
- [3] J. Dénes, A.D. Keedwell, *Latin Squares*, Elsevier, New York, 1991.
- [4] D.A. Drake, Maximal sets, latin squares and partial transversals, *J. Statist. Plann. Inference* 1 (1977) 143–149.
- [5] P. Erdős, J. Spencer, Lopsided Lovász local lemma and latin transversals, *Discrete Appl. Math.* 30 (1991) 151–154.
- [6] L. Euler, (Leonhard Euleri), *Opera Omnia* 7 (1923) 391–392.
- [7] H.-L. Fu, S.-C. Lin, C.-M. Fu, The length of a partial transversal in a latin square, *J. Comb. Math. Combin. Comput.* 43 (2002) 57–64.
- [8] M. Hall Jr., A combinatorial problem on groups, *Proc. Amer. Math. Soc.* 3 (1952) 584–587.
- [9] D.R. Hickerson, personal communication, 2003.
- [10] K.K. Koksma, A lower bound for the order of a partial transversal in a latin square, *J. Combin. Theory Ser. A* 7 (1969) 94–99.
- [11] H.W. Norton, The 7×7 squares, *Ann. Eugenics* 9 (1939) 269–307.
- [12] L.J. Paige, A note on finite abelian groups, *Bull. Amer. Math. Soc.* 53 (1947) 590–593.
- [13] E.T. Parker, correspondence, 1989.
- [14] H.J. Ryser, Neuere Problem in der Kombinatorik, in *Vorträge über Kombinatorik*, Oberwolfach (1967) 69–91.
- [15] P.W. Shor, A lower bound for the length of partial a transversal in a latin square, *J. Combin. Theory Ser. A* 33 (1982) 1–8.
- [16] H.S. Snevily, The Cayley addition table of Z_n , *Amer. Math. Monthly* 106 (1999) 584–585.
- [17] S.K. Stein, Transversals of latin squares and their generalizations, *Pacific J. Math.* 59 (1975) 567–575.
- [18] J. van Rees, correspondence, 2005.
- [19] D.E. Woolbright, An $n \times n$ latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols, *J. Combin. Theory Ser. A* 24 (1978) 235–237.